

# Learning L2 pronunciation with a text-to-speech synthesizer

Walcir Cardoso<sup>1</sup>

**Abstract.** The study explored the pedagogical use of Text-To-Speech (TTS) synthesizers by comparing two groups of students learning the pronunciations associated with Regular Past Tense (RPT) marking in English (e.g. talk[t], clean[d], want[id]). While one group received TTS-based instruction, the other was taught RPT with teacher assistance. Our findings indicate that both groups behaved similarly regarding the production of RPT, and that the target allomorphs are acquired in the following order: *id>d>t*.

**Keywords:** text-to-speech synthesis, L2 pronunciation.

## 1. Introduction

TTS are applications that convert written text into spoken voice output. Although the literature on its pedagogical applications in L2 education is still scarce, the few available studies indicate that the technology has potential for the teaching/learning of L2 pronunciation (Liakin, Cardoso, & Liakina, 2017). Research has also shown that the technology is ready for use in the L2 classroom, as its output is comparable to that of humans on measures such as intelligibility, comprehensibility, and morphophonemic accuracy (Cardoso, Smith, & Garcia-Fuentes, 2015).

We hypothesize that TTS can contribute to the learning of L2 pronunciation because it can enhance the target input in both quantity and quality via features such as speed control (slow/fast speech), multimodal visual presentation (through orthography and the highlighting of words being read/synthesized), word/phrase repetition, and voice selection (e.g. male or female voices for high phonetic

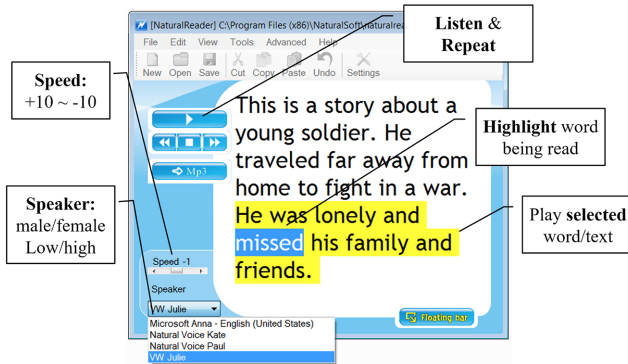
---

1. Concordia University, Montreal, Canada; walcir.cardoso@concordia.ca

**How to cite this article:** Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouéšny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 16-21). Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.26.806>

variability, as recommended by Logan, Lively, & Pisoni, 1991). These features are shown in Figure 1.

Figure 1. Input enhancement via TTS: features (reproduced with permission from NaturalSoft)



To examine the pedagogical suitability of TTS, this study focused on the acquisition of the allomorphy found in RPT marking in English (i.e. talk[t], clean[d] and want[id]). As suggested in the literature (Collins et al., 2009), the acquisition of RPT and related morphophonemics is difficult because -ed forms are not readily available to L2 learners (an issue of input ‘quantity’) and they occur in hard-to-perceive phonetic contexts (an issue of input ‘quality’). By allowing learners to manipulate the quantity and quality of exposure to the forms being learned, TTS can address the issues of availability and accessibility reported in Collins et al. (2009). This study was guided by the following question:

- Which type of instruction is more effective in the learning of simple past -ed allomorphy: TTS-based instruction or one led by a language teacher?

## 2. Method

### 2.1. Participants and experimental groups


Eighteen English-as-a-foreign-language students (proficiency: intermediate; age: 13-22; gender: seven females, 11 males) were stratified among two groups: while the TTS Group was asked to complete learning activities using the TTS application in Figure 1 (thus manipulating the L2 input in quantity and quality), the Non-

TTS Group received the same treatment with the assistance of an English teacher. Over a four-week period, all participants engaged in incidental learning of RPT allomorphy via listening activities that focused on the pronunciation of the target forms (e.g. listen and fill-in the gap, and sound categorization).

2.2. Instruments

The data were collected through two oral tasks consisting of 70 ed-final verbs and 30 distractors: Word Reading (id=10; d=23; t=12) and Guided Task (id=3; d=14; t=8) in which participants were asked to answer questions about the past holidays of fictional characters (see Figure 2 for an illustration).

Figure 2. Guided oral task



Last July, Lena took a four-week vacation in Japan. What did she do?

	YES	NO
arrive safely?	✓	
work during her vacation?	✓	
call her boss every day?		✓ (week)
stay at a hotel?		✓ (with friends)
wash her clothes every day?	✓	
clean her bedroom?	✓	
dress informally?		✓ (formally)
need guides to travel in Japan?	✓	
love the weather?	✓	
travel with her friends?		✓ (boyfriend)

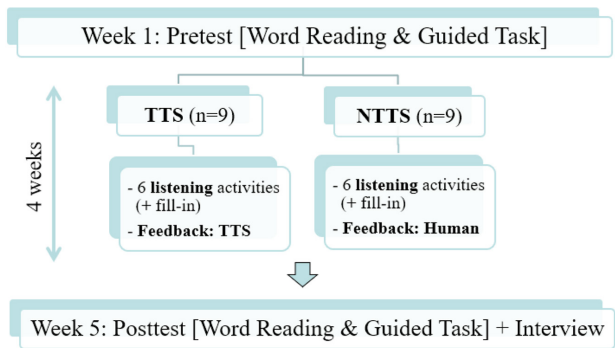
2.3. Research design and procedures

The study followed a pretest-posttest design. Participants were pretested using the two instruments described above and tested within one week of completion of the learning activities described above. The posttest was followed by a 20-minute oral interview in which participants were asked about their pedagogical experience. Figure 3 illustrates the research design and procedures adopted.

The data were coded by two research assistants (1=correct, 0=incorrect) and checked for consistency by a third rater (95.56% rater consistency). The interview

data were audio-recorded, transcribed, and categorized into pertinent themes (e.g. morphophonemic awareness and perceived pedagogical benefits).

Figure 3. Research design and procedures



3. Results

Due to space limitations, only the final stage of the statistical analysis is reported in detail. Mixed between-within subjects analysis of variance (ANOVA), followed by pairwise comparisons, were conducted with RPT allomorphy (-id, -d, -t) as dependent variable, and Time (pretest, posttest), Group (TTS, Non-TTS), and Task (Word Reading, Guided Task) as independent variables. For both Word Reading and Guided Task, there was no effect for Time or Group, indicating no significant improvement: the RPT allomorphs remained the same from pretest to posttest, across the two groups and tasks. There was, however, a difference for RPT production in the two tasks on the pretests and posttests, attesting that the three allomorphs behaved significantly differently; while -id was unproblematic, -d was relatively less difficult to produce than -t: id>d>t (where > indicates “easier to produce than” or “acquired before”). Finally, additional analysis revealed no significant improvement in Word Reading. Due to the uneven distribution of RPT forms, the results are provided in mean percentages in Table 1.

Table 1. RTP allomorphs by group (type of instruction), time, and task

Group	Pretest (%)						Posttest (%)					
	Word Reading			Guided Task			Word Reading			Guided Task		
	-id	-d	-t	-id	-d	-t	-id	-d	-t	-id	-d	-t
TTS	100	20	8.9	75	30.6	13.9	100	18.8	7.41	63	42.9	13.9
Non-TTS	97.8	13.3	2.2	66.7	19.4	8.3	95.6	10.1	0	74.1	23	6.9

Because the two groups were statistically similar and no improvement was observed in Word Reading, follow-up Within-Subjects ANOVAs involving Guided Task and the two groups combined were conducted. This revealed an interaction between Time and RTP allomorphy ( $F(1,17)=15.609$ ,  $p=0.001$  (alpha level set at .003), suggesting a performance improvement for -d (shaded in Table 1), but not for -t on the posttest.

To summarize, the TTS and Non-TTS groups behaved similarly regarding the production of RPT: while they both improved in producing [d], there was no significant improvement in their production of -id (participants scored at the ceiling level on the pretest) and [t].

## 4. Discussion and concluding remarks

The study examined the effects of two types of instruction in the learning of RPT allomorphy: TTS-based instruction and one led by a language teacher. The results indicate that the TTS and Non-TTS groups behaved similarly regarding the production of RPT. An interesting pattern emerged in the analysis: an improvement of allomorph -d by both groups, suggesting a development sequence in the acquisition of RPT allomorphy. While -ed is acquired first, due to L1 influence (see Cardoso, 2011), the order -d>-t can be explained by the fact that syllabic coda -d is more sonorous than -t (languages favor sonorous codas; Cardoso, 2011).

The lack of major improvements observed can be attributed to at least two factors. Firstly, it is possible that an exclusive focus on perception via the proposed TTS-based listening activities (under the ‘perception precedes production’ hypothesis – Flege, 1995) was not appropriate to lead learners to transfer the newly-acquired knowledge into production. If these activities had been complemented with speaking activities (e.g. via speech recognition software or even silent practice), learners would have had an opportunity to practice and possibly transfer this novel perceptual knowledge into production. That many learners became aware of RPT allomorphy became evident in the analysis of their interview data, particularly among participants in the TTS Group. When asked what they had learned in the study, most participants reported that they had become aware of the differences in RPT pronunciation: “if I were to [listen to these stories] again, I wouldn’t pronounce as much this “ed” at the end”; “I never knew that the {ed} sometimes sounds strong, sometimes weak”. Secondly, it is possible that the four hours allotted for listening practice was not enough, particularly in an experiment in which

pronunciation was learned implicitly (see [Saito & Lyster, 2012](#), for the effects of explicit pronunciation teaching).

An important pedagogical implication of these findings is that TTS has the potential to increase and improve the target language input (preferably accompanied by explicit instruction and oral practice), develop better grapheme-to-phoneme associations (there is no clear relationship between orthographic -ed and its allomorphs -id, -d and -t in English), and address some of the time constraints that affect many L2 learning environments (TTS has the potential to foster autonomous and anytime-anywhere learning). Using TTS as an out-of-class pedagogical tool could then increase in-class time so that teachers and students could focus on other important tasks such as providing feedback and engaging in real-life communicative activities.

## References

- Cardoso, W. (2011). The development of coda perception in second language phonology: a variationist perspective. *Second Language Research*, 27(4), 433-465.
- Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesizers. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp.108-113). Research-publishing.net. <https://doi.org/10.14705/rpnet.2015.000318>
- Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question. *Modern Language Journal*, 93(3), 336-353. <https://doi.org/10.1111/j.1540-4781.2009.00894.x>
- Flege, J. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233-72). York Press.
- Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis: focus on French liaison. *Computer Assisted Language Learning*, 30(3-4), 348-365. <https://doi.org/10.1080/09588221.2017.1312463>
- Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*, 89(2), 874-886.
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning*, 62(2), 595-633. <https://doi.org/10.1111/j.1467-9922.2011.00639.x>

Published by Research-publishing.net, a not-for-profit association  
Contact: [info@research-publishing.net](mailto:info@research-publishing.net)

© 2018 by Editors (collective work)  
© 2018 by Authors (individual work)

**Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018**  
Edited by Peppi Taalas, Juha Jalkanen, Linda Bradley, and Sylvie Thouéšny

**Publication date: 2018/12/08**

**Rights:** the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2018.26.9782490057221>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

**Disclaimer:** Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

**Trademark notice:** product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Copyrighted material:** every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net  
Cover theme by © 2018 Antti Myöhänen ([antti.myohanen@gmail.com](mailto:antti.myohanen@gmail.com))  
Cover layout by © 2018 Raphaël Savina ([raphael@savina.net](mailto:raphael@savina.net))  
Drawings by © 2018 Linda Saukko-Rauta ([linda@redanredan.fi](mailto:linda@redanredan.fi))

ISBN13: 978-2-490057-22-1 (Ebook, PDF, colour)

ISBN13: 978-2-490057-23-8 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-21-4 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.  
A cataloguing record for this book is available from the British Library.

**Legal deposit, UK:** British Library.

**Legal deposit, France:** Bibliothèque Nationale de France - Dépôt légal: Décembre 2018.